Current Microbiology Vol. 50 (2005), pp. 78–83
DOI: 10.1007/s00284-004-4412-6

Current Microbiology
An International Journal

Genetic Discovery in *Xylella fastidiosa* Through Sequence Analysis of Selected Randomly Amplified Polymorphic DNAs

Jianchi Chen,[1] Edwin L. Civerolo,[1] Robert L. Jarret,[2] Marie-Anne Van Sluys,[3] Mariana C. de Oliveira[3]

[1]Agricultural Research Service, U. S. Department of Agriculture, Parlier, CA 93648, USA
[2]Agricultural Research Service, U. S. Department of Agriculture, Griffin, GA 30223, USA
[3]Instituto de Biociências, Universidade de São Paulo, Caixa Postal 11461, cep 05422-970 São Paulo, SP, Brazil

Accepted: 13 September 2004

Abstract. *Xylella fastidiosa* causes many important plant diseases including Pierce's disease (PD) in grape and almond leaf scorch disease (ALSD). DNA-based methodologies, such as randomly amplified polymorphic DNA (RAPD) analysis, have been playing key roles in genetic information collection of the bacterium. This study further analyzed the nucleotide sequences of selected RAPDs from *X. fastidiosa* strains in conjunction with the available genome sequence databases and unveiled several previously unknown novel genetic traits. These include a sequence highly similar to those in the phage family of Podoviridae. Genome comparisons among *X. fastidiosa* strains suggested that the ''phage'' is currently active. Two other RAPDs were also related to horizontal gene transfer: one was part of a broadly distributed cryptic plasmid and the other was associated with conjugal transfer. One RAPD inferred a genomic rearrangement event among *X. fastidiosa* PD strains and another identified a single nucleotide polymorphism of evolutionary value.

Xylella fastidiosa causes diseases on many economically important plants. In California, Pierce's disease (PD) of grapevine is currently threatening grape production in several areas and almond leaf scorch disease (ALSD) has re-emerged in the past few years. The potential destructive impact of *X. fastidiosa* led to substantial efforts to sequence the bacterial genomes of four *X. fastidiosa* strains. The genome sequences of *X. fastidiosa* strain Temecula ATCC700964, causing PD, and *X. fastidiosa* strain 9a5c (citrus variegated chlorosis, CVC) have been completed [19, 21]. The sequences of *X. fastdiosa* strain Dixon (almond leaf scorch disease, ALSD) and *X. fastidiosa* strain Ann-1 (oleander leaf scorch disease, OLSD) remain gapped [3]. Comparisons of these genome sequences revealed intriguing genetic and biological insights about the bacterium. Ninety-eight percent of the PD-Temecula genes are shared with those in the CVC-9a5c genome. Horizontal gene transfer and chromosomal rearrangements contributed significantly to genomic differentiation. Horizontally acquired elements, such as prophages, plasmids, and genomic islands, contributed up to 18% of the final genome indicative of *X. fastidiosa*'s highly flexible gene pool [14]. A limited number of strains, however, do not contain all the genetic material and the encoded information of a species. During the process of environmental adaptation such as host specialization and pathogenesis development, a bacterial genome could be subject to base mutation, sequence deletion and/or insertion (indel), horizontal gene transfer, and genome rearrangement. Such genetic information can only be acquired through genomic comparison involving other strains from different sources. Prior to the availability of genome sequence databases, genome comparisons were achieved using random or presumptively random sampling approaches such as analysis of randomly amplified polymorphic DNA (RAPDs) or restriction fragment length polymorphisms (RFLPs) [5, 6, 15, 18]. These types of analyses compared similarities and differences among PCR-generated genomic DNA fragments flanked by defined 10-base sequences, or the presence or absence of restriction enzyme sites.

Correspondence to: Jianchi Chen; *email:* jichen@fresno.ars.usda.gov

Table 1. List of strain-unique RAPDs/DNA sequences that reveal novel genetic traits and the related information

RAPD/sequence name	Size (bp)	Primer/flanking sequence	Putative genetic information	New genetic trait	Accession number
PD1-1-2	1005	5′-TCAGGCCCTT-3′	ORF, $tRNA^{ser}$	Conserved SNP	AF130454
PDX1-2	661	5′-GTGATCGCAG-3′	Ubiquinol cytochrome C oxidoreductase	Genomic rearrangement	AY270178
PDX3-1	547	5′-TGCCGAGCTG-3′	Integrase	Phage	AY270180
PDX6-1	966	5′-GAAACGGGTG-3′	Replicase, ORF	Plasmid	AY457972
M23#18-33	493	5′-AGGTGACCGT-3′	Conjugal protein	Type IV conjugation	AY665687
Dixon-037	555	5′-TGCCGAGCTG-3′	Integrase	Phage	NZ_AAAL01000037
Dixon-072r	547	5′-TGCCGAGCTG-3′	Integrase	Phage	NZ_AAAL01000072
OLSD-Ann-1r	547	5′-TGCCGAGCTG-3′	Integrase	Phage	NZ_AAAM01000074

Fragments analyzed were generally approximately 0.1–21 kb within the resolution size limit of agarose gel electrophoresis. The electrophoretic profiles of DNA fragments based on size were scored in a binary format and variations were summarized by statistical means. Because of its technical simplicity, RAPD analysis is a common tool for genetic comparison among bacterial strains. However, the total genetic information embedded in the RAPD fragments is not utilized.

Albibi et al. [1] evaluated the use of RAPDs to differentiate *X. fastidiosa* strains of different host origins. One RAPD was subsequently sequenced and used to develop PCR primers specific for a PD strain [2]. Sequence analysis using the then available databases indicated that the PD strain-specific site was located in an intergenic region. Such sequence analyses revealed genetic information more than the primer site and DNA fragment length [2, 16]. In this project, we hypothesized that at least some size-unique RAPDs contain novel genetic information currently unknown. RAPDs were selected through strain comparisons and the sequences were used to search for new genetic information. A particular focus was on PD-related *X. fastidiosa* strains.

Materials and Methods

Bacterial strains and DNA sequences. Two *X. fastidiosa* strains were used to generate RAPDs and obtain nucleotide sequences. A field strain, designated as PD28-5, was isolated from a Florida bunch grape cultivar, ''Orlando Seedless,'' a hybrid of *Vitis vinifera* and other *Vitis* species native to the southeastern United States. Another PD-related strain, M23, was isolated from an almond cultivar, ''Sonora,'' in California. DNA was extracted from the bacterial strains and DNA fragments were amplified using PCR as previously described [1]. RAPD profiles were compared to those of *X. fastidiosa* strains including almond leaf scorch, periwinkle wilt, phony peach, plum leaf scald, and oak leaf scorch to identify size-unique RAPDs. For PD28-5, the unique DNAs were isolated from agarose gels and re-amplified following the procedure previously described [4]. The amplified fragments were cloned and sequenced as described by Banks et al. [2]. For M23, total PCR products were cloned. The cloned DNA fragments were re-correlated to the correspondent DNA bands by size and sequenced.

Genome sequence database searching. BLASTN (for DNA sequence) and BLASTX (for putative amino acid sequence) were used for sequence similarity searches on the BLAST network service at the National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov). Fragment sequences were first used to search against the non-redundant sequence databases to check for specificity, and then limited to the genome sequences of four *X. fastidiosa* strains: PD-Temecula, ALSD-Dixon, and OLSD-Ann-1 from California, and CVC- 9a5c from Brazil. Top hit sequences were evaluated according to statistics provided by the program and the length of DNA segments selected by default according to the BLAST algorithm. Putative genes were assigned based on sequence information from the annotated genomes.

Comparison of *X. fastidiosa* strains. For pairwise comparison, the Pairwise BLAST program from the NCBI network service was used. For multiple sequence comparisons, sequences of DNA homologs were retrieved from the GenBank database and aligned using CLUSTAL-W [20] through the network service of the European Bioinformatics Institute (http://www.ebi.ac.uk).

Pathogenicity test. *X. fastidiosa* Temecula was cultured on PW medium solidified by Gel-Rite [10] at 28°C for 10 days. Bacterial cells were scraped from the medium and suspended in distilled water. The suspension was adjusted to 0.2 OD at 620 nm. Ten microliters of cell suspension were inoculated into grape cultivar Cabernet Sauvignon by the pin-prick method [12] in a greenhouse. Pathogenicity was indicated by the presence of marginal necrosis symptom and isolation of *X. fastidiosa* cells from the petioles one month after inoculation.

Nucleotide sequence accession numbers. Nucleotide sequence data have been deposited in GenBank and are listed in Table 1.

Results and Discussion

A total of 14 RAPDs (seven from PD28-5 and seven from M23), ranging from 200 to 1200 bp and unique to at least one non-PD strain, were selected and sequenced. BLASTN searches against the GenBank database indicated that all of the RAPD sequences were most similar to the four *X. fastidiosa* genome sequences over all of the available bacterial DNA sequences deposited in

GenBank. With the exception of PDX3-1 and M23-A18-33 (the former was from a putative phage and the latter from a plasmid), the 12 DNA sequences were most similar to PD-Temecula. These results illustrate the genome congruency of *X. fastidiosa* at the species level and possibly at the pathotype level. While Chen et al. [7, 8] reported that short 16S rDNA sequences could be used to identify *X. fastidiosa*, this study further suggests that randomly selected DNA sequences also have the capacity to be species specific.

Whole sequence analyses of nine out of the 14 RAPDs did not reveal novel genetic information beyond the description in the four *X. fastidiosa* genome sequence annotations. The size uniqueness of these sequences could be attributed to the single nucleotide polymorphisms (SNPs) at the primer priming sites because SNPs were found at both 5′ and 3′ ends. However, SNPs were also common along the main DNA sequences ranging from 4 to 97 depending on a particular sequence. The other five RAPD sequences, listed in Table 1, revealed significant new genetic information. Three (PDX3-1, PDX6-1, M23-A18-33) were indicative of horizontal transfer, one (PDX1-2) was suggestive of a genome rearrangement, and one (PD1-1-2) identified a SNP responsible for the annotation difference between the genomes of PD-Temecula and CVC-9a5c. The major genetic findings are discussed below.

An active "phage" in the *X. fastidiosa* population. The homolog of PDX3-1 was not present in either the PD-Temecula or CVC-9a5c strain. However, homologs were present in both ALSD-Dixon and OLSD-Ann-1 strains. Considering the close rela- tionship among *X. fastidiosa* strains, the occurrence of this homolog can be taken as evidence for a mobile genetic element such as a phage. Sequence analysis substantiated this conclusion. The sequence of PDX3-1 contained three sections: (1) a C-terminal portion of an integrase gene *int*; (2) a spacer sequence of 16 bp downstream of *int*; and (3) a 43-bp sequence at the 3′ end, which was almost identical (42/43 bp) to that of the anti-codon and TΨC arms of a $tRNA^{lys}$ gene found in all *X. fastidiosa* genomes.

The int amino acid sequence of PDX3-1 shared a high degree of homology to int amino acid sequences of several enterobacterial phages. Among these bacteriophages were ST64T from *Salmonella typhimurium* (47% identities, 64% positives), bacteriophage V from *Shigella flexneri* (45% identities, 64% positive), and bacteriophage APSE-1 from a symbiont of an aphid (*Acyrthosiphon pisum*) (44% identities, 61% positive). All of these enterobacterial phages belong to the family Podoviridae and have a double-stranded DNA genome. The $tRNA^{lys}$-like sequence could be the integrase attachment site (*att*). Partial or full tRNA genes are commonly used as integration sites for many phage-related integrases [22].

Because of the availability of the whole genome sequence, strain PD-Temecula is a common reference in *X. fastidiosa* research. During the preparation of genome sequencing, the bacterial strain must have been subjected to extensive sub-culturing, a process that could lead to the loss of virulence [11]. The lack of PDX3-1 "phage" in PD-Temecula raised a question if the "phage" was related to virulence. However, results of our pathogenicity test indicated that PD-Temecula was highly pathogenic as evidenced by the presence of typical marginal necrosis symptoms, re-isolation of the bacterium, and high mortality rate of the inoculated hosts (5/5 within one year). Therefore, the PDX3-1 "phage" does not harbor gene(s) directly responsible for virulence.

It was of interest to identify two copies of a PDX3-1 homolog in the ALSD-Dixon genome. One of these homologs was found in contig NZ_AAAL01000072.1 and the other in contig NZ_AAAL01000037.1. In the latter contig, the spacer sequence between *int* and *att* was highly heterologous and included an 8-bp insertion (5′ tctttctg 3′). In the genome sequence of OLSD-Ann-1, one PDX3-1 homolog, located in contig NZ_AAAM01000074.1, was found. Bhattacharyya et al. [3] analyzed the draft genome sequences of the ALSD-Dixon and OLSD-Ann-1 strains and the complete genome sequence of the CVC-9a5c strain. They identified a 65-kb phage insertion region unique to the CVC-9a5c strain, but did not address the presence of PDX3-1 homologs in the ALSD-Dixon and OLSD-Ann-1 strains and its absence in the CVC-9a5c strain. However, distantly related putative proteins belonging to the same clusters of the orthologous group, COG0582, as the int of PDX3-1 does, were present in the CVC-9a5c genome.

A cryptic plasmid family. PDX6-1 was found to be highly similar to three small plasmids, pXFPD1.3 of PD-Temecula, pXF13 of CVC-9a5c, and pXF868 from a mulberry leaf scorch strain (ATCC35868) [17]. No homolog was found in the genome of the ALSD-Dixon or OLSD-Ann-1 strains. Because of the gapped nature of the genome databases of these strains, it may be premature to conclude that the PDX6-1 homolog is absent in both genomes. However, Bhattacharyya et al. [3] did not report the presence of plasmids in the two genomes. Hendson et al. [9] reported the occurrence of plasmids in both the ALSD-Dixon and OLSD-Ann-1 strains. However, plasmid fragments of 1.3 kb or smaller in size were not reported.

X. fastidiosa plasmids were first reported by Chen et al. [4]. A plasmid about 1.3 kb was observed in PD and periwinkle wilt strains from Florida and a mulberry leaf scorch strain from Massachusetts. The latter was sequenced by Pooler et al. [17]. DNA hybridization data indicated that the plasmids from grape, periwinkle, and mulberry were homologous [4]. Sequence alignments of PDX6-1 to pXFPD1.3, pXF13, and pXF868 showed that these plasmids were highly co-linear with moderate sequence similarity. The best alignment from BLASTN search ranged from 79% to 86% for DNA segment over 500 bp. Sequence annotation suggested the presence of two open reading frames, one for a replicase gene and the other unknown. The plasmid is not essential for bacterial replication because it is not present in many other strains of *X. fastidiosa* [4]. The plasmid is small but widely distributed occurring from Florida to California and from Massachusetts to Brazil. Discounting the replicase gene, the plasmid genome is questionably large enough to encode an additional gene. The biological function of such a small self-replicated element remains to be determined.

A plasmid related gene for conjugal transfer. M23-A18-33 is highly similar to part of the gene XfasA0256 in the ALS-Dixon genome. The two sequences were almost identical (99%, 465/467 bp). The translated amino acid sequences were also highly similar (99%, identities = 153/154). The product of XfasA0256 is annotated as VirB4, a component of Type IV secretory pathway, responsible for conjugal DNA transfer. The M23-A18-33 ortholog was also found in the OLSD Ann-1 genome with a 96% or 449/467 bp similarity. M23-A18-33 also shared a high degree of sequence similarity (96% or 450/467 bp) to a section of a plasmid, pXF51, identified during the whole genome sequencing of CVC-9a5c [19]. A similar sequence was also present in the CVC-9a5c chromosome, which was believed originate from the pXF51 [13]. Whether M23-A18-33 resides in a plasmid or the bacterial chromosome remains unknown. However, the almost identical nature between M23-A18-33 and XfasA0256 indicates the horizontal gene transfer among the two different genotypes of *X. fastidiosa* isolated from almond in California. On the other hand, although strains PD-Temecula and ALS-M23 are genetically very similar, PD-Temecula does not harbor the M23-A18-33 homolog (Fig. 1), a further indication of the mobility of M23-A18-33 element.

A genomic rearrangement in PD strains. PDX1-2 has a sequence length of 661 bp. The sequence encodes a portion of a cytochrome C oxidoreductase (Table 1) belonging to COG1290. Homologs of PDX1-2 are present in the four *X. fastidiosa* genomes but are split and located in two distinct locations. The insertion is 126 kb in the PD-Temecula genome and 468 kb in the CVC-9a5c genome. The insertion in the draft genome sequences of the ALSD-Dixon and OLSD-Ann-1 strains could not be estimated due to the involvement of different contigs. In the genome of the PD-Temecula strain, the insertion starts from the intergenic region between ORF1671 and ORF 1672 and ends in ORF1776. This covered 104 ORFs. The insertion size in the genome of CVC-9a5c strain is three times longer than that in the PD-Temecula strain and spanned from ORF0400 to ORF0909, or 509 ORFs. Further analysis indicated that the insertions are part of a genome rearrangement

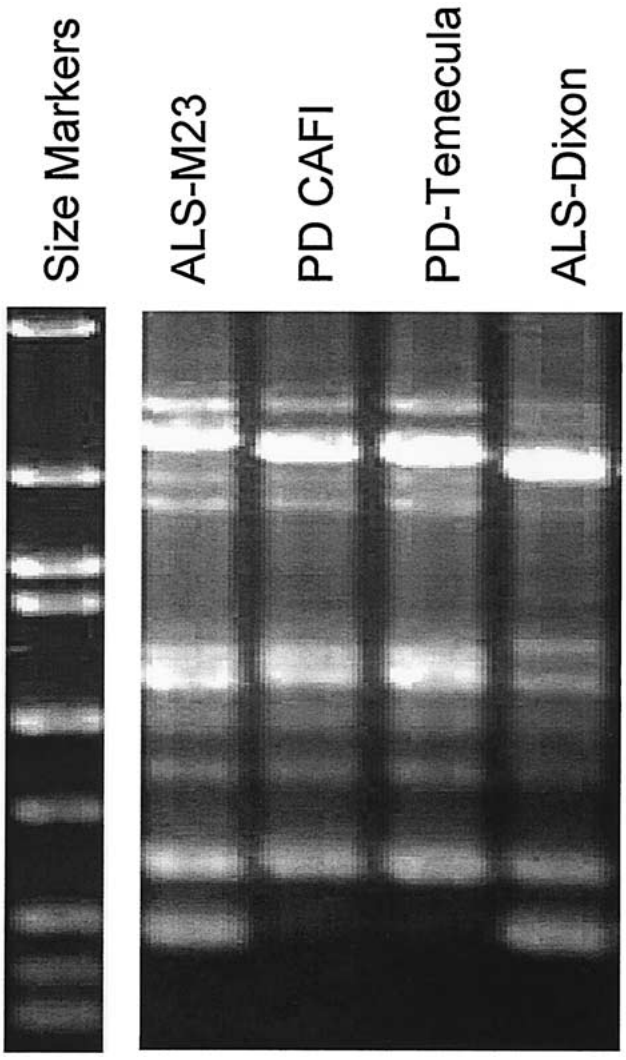

Fig. 1. A RAPD profile showing the size-unique DNA band, M23-A18-33 (arrow). Size marker (from bottom) in bp: 300, 400, 500, 750, 1000, 1400, 1550, 2000, and 3000. ALS-M23, PD CAFI, and PD-Temecula are grape PD related, while ALS-Dixon is not.

differentiating the PD-Temecula and CVC-9a5c strains [2]. Because of the large number of ORFs, the absence of the DNA section in the Florida PD strain is likely due to a genomic DNA rearrangement rather than a simple deletion. Further supporting this hypothesis is the presence of a pair of homologous (84%) direct repeats, upstream 36 bp (5′ taaatcggtt ggtgtgttag tgatgttttc tgcgat 3′) and downstream 39 bp (5′ taaatctcgt tgttcgtgtt cgatgttgtt ctcagcgat 3′, i.e., three insertions and six SNPs). Sequence repeats are known to be related to chromosomal insertion/deletion events. Genome rearrangements are characteristic of the CVC-9a5c and PD-Temecula strains [2]. Our data suggest that genome rearrangements also occur within PD strains, adding further evidence to support the suggestion that the genome of *X. fastidiosa* is highly plastic [4].

A SNP of evolutionary value. A portion of PD1-1-2 was homologous with ORF XF0126, of the CVC-9a5c genome. The homolog is, however, not annotated in the PD-Temecula, ALSD-Dixon, or OLSD-Ann-1 genomes. Sequence analysis revealed that the 27th amino acid of XF0126 from CVC-9a5c was TGG, coding for Trp. In the genomes of the PD-Temecula, ALSD-Dixon, and OLSD-Ann-1 strains, the third position of the corresponding codon was a transitional mutation from G to A. This resulted in a termination codon TGA that led to a putatively truncated peptide. This truncated peptide was not annotated in any of the three genome sequences. The Florida PD strain contained the same transitional mutation. Based on the genome sequence database, this transitional mutation, also regarded as SNP, separated the South American CVC-9a5c from the North American *X. fastidiosa* strains. Such a SNP will be useful for strain differentiations.

Another feature of PD1-1-2 is the physical location of its homologs. In both the PD-Temecula and CVC-9a5c chromosomes, these homologs are located between the two *rrn* operons towards the putative replication sites. This portion of the genome is not subjected to genome rearrangement [21] and is possibly more ancient and conserved. Therefore, the SNP in XF0126 could serve as a point marking the separation of the South American and North American *X. fastidiosa* strains.

In summary, this study utilizes the information from RAPD analyses in conjunction with the available genome sequence database to identify a new ''phage'' element, a widely distributed plasmid family, a conjugal gene, a genomic rearrangement within PD strain, and a SNP of evolutionary importance. The results support our research hypothesis that some size-unique RAPDs contain novel genetic information (5/14 or 36%), extend our current knowledge about the population genomics of *X. fastidiosa*, and demonstrate an additional method for comparative genomic research. It should be noted that although whole genome sequence analyses [3, 21] have revealed the high incidence of phage-related sequences in the *X. fastidiosa* genome, this study provided the first physical evidence for the presence an active ''phage.''

Literature Cited

1. Albibi R, Chen J, Lamikanra O, Banks D, Jarret RL, Smith BJ (1998) RAPD fingerprinting *Xylella fastidiosa* Pierce's disease strains isolated from a vineyard in North Florida. FEMS Microbiol Lett 165:347–352
2. Banks D, Albibi R, Chen J, Lamikanra O, Jarret RL, Smith BJ (1999) Specific detection of *Xylella fastidiosa* Pierce's disease strains. Curr Microbiol 39:85–88
3. Bhattacharyya A, Stilwagen S, Reznik G, Feil H, Feil WS, Anderson I, Bernal A, D'Souza M, Ivanova N, Kapatral V, Larsen N, Los T, Lykidis A, Selkov E Jr, Walunas TL, Purcell A, Edwards RA, Hawkins T, Haselkorn R, Overbeek R, Kyrpides NC, Predki PF (2002) Draft sequencing and comparative genomics of *Xylella fastidiosa* strains reveal novel biological insights. Genome Res 12:1556–63
4. Chen J, Chang CJ, Jarret RL (1992) Plasmids from *Xylella fastidiosa* strains. Can J Microbiol 38:993–995
5. Chen J, Chang CJ, Jarret RL, Gawel N (1992) Genetic variation among *Xylella fastidiosa* strains. Phytopathology 82:973–977
6. Chen J, Lamikanra O, Chang CJ, Hopkins DL (1995) Randomly amplified polymorphic DNA analysis of *Xylella fastidiosa* Pierce's disease and oak leaf scorch pathotypes. Appl Environ Microbiol 61:1688–1690
7. Chen J, Banks D, Jarret RL, Chang CJ, Smith BJ (2000) Use of 16S rDNA sequences as signature characters to identify *Xylella fastidiosa*. Curr Microbiol 40:29–33
8. Chen J, Jarret R, Qin X, Hartung JS, Banks D, Chang CJ, Hopkins DL (2000) 16S rDNA analysis of *Xylella fastidiosa* strains. Syst Appl Microbiol 23:349–354
9. Hendson M, Purcell AH, Chen D, Smart C, Guilhabert M, Kirkpatrick B (2001) Genetic diversity of Pierce's disease strains and other pathotypes of *Xylella fastidiosa*. Appl Environ Microbiol 67:895–903
10. Hill BL, Purcell AH (1995) Acquisition and retention of *Xylella fastidiosa* by an efficient vector, *Graphocephala atropunctata*. Phytopathology 85:209–212
11. Hopkins DL (1989) *Xylella fastidiosa*: Xylem-limited bacterial pathogen of plants. Annu Rev Phytopathol 27:271–290
12. Hopkins D (2001) *Xylella fastidiosa*. In: Schaad NW, Jones JB, Chun W (eds) Laboratory guide for identification of plant pathogenic bacteria. St. Paul, MN: APS Press, pp 201–213
13. Marques MV, da Silva AM, Gomes SL (2001) Genetic organization of plasmid pXF51 from the plant pathogen *Xylella fastidiosa*. Plasmid 45:184–199
14. Nunes LR, Rosato YB, Muto NH, Yanai GM, da Silva VS, Leite DB, Goncalves ER, de Souza AA, Coletta-Filho HD, Machado MA, Lopes SA, de Oliveira RC (2003) Microarray analyses of *Xylella fastidiosa* provide evidence of coordinated transcription control of laterally transferred elements. Genome Res 13:570–578
15. Pooler MR, Hartung JS (1995) Genetic relationships among strains of *Xylella fastidiosa* from RAPD-PCR data. Curr Microbiol 31:134–137

16. Pooler MR, Hartung JS (1995) Specific PCR detection and identification of *Xylella fastidiosa* strains causing citrus variegated chlorosis. Curr Microbiol 31:377–381
17. Pooler MR, Hartung JS, Fenton RG (1997) Sequence analysis of a 1296-nucleotide plasmid from *Xylella fastidiosa*. FEMS Microbiol Lett 155:217–222
18. Rosato YB, Neto JR, Miranda VS, Carlos EF, Manfio GP (1998) Diversity of a *Xylella fastidiosa* population isolated from *Citrus sinensis* affected by citrus variegated chlorosis in Brazil. Syst Appl Microbiol 21:593–598
19. Simpson AJG, Reinach FC, Arruda P (2000) The genome sequence of the plant pathogen *Xylella fastidiosa*. Nature 406:151–157
20. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680
21. Van Sluys MA, de Oliveira MC, Monteiro-Vitorello CB, et al. (2003) Comparative analyses of the complete genome sequences of Pierce's disease and citrus variegated chlorosis strains of *Xylella fastidiosa*. J Bacteriol 185:1018–1026
22. Williams KP (2002) Survey and summary: Integration sites for genetic elements in prokaryotic tRNA and tmRNA genes: sublocation preference of integrase subfamilies. Nucleic Acids Res 30:866–875